MSPA PRED 420
Session 1 Python Practice
Getting Started
v.5

5

This course is about examining and manipulating data using Python.   Throughout the course
you'll be doing "Practices," sets of exercises that illustrate key concepts in using Python for
what is sometimes called "data munging." Note that this course and its exercises assume
some prior experience with Python.

10

The data you'll be using are from a firm we'll call the XYZ company. The data are real. The
firm's identity is concealed to protect the guilty. More about XYZ shortly. But first here's a bit
to get you set up to do your exercises in Python.

15   Different Python environments are available. The one that the MSPA uses in its courses is
the Canopy distribution from Enthought, which uses Python 2.7. There are different versions
of Canopy. The academic version can be downloaded for free from Enthought if you register
with them on their website (https://enthought.com/) using your Northwestern email address. If
you haven't yet downloaded and installed Canopy, you should do so now.   Versions that run
20   on Windows, Mac, and Linux are available. You should use a 64 bit version if your hardwared
will support it.

You'll find that installing Canopy is pretty straightforward. Once you've installed it you should
explore it a little by poking around in it. If you installed it in it's default form, when you launch
25   it it will show you a "Welcome to Canopy" window that has has options like "Editor," "Package
Manager," and "Doc Browser" The Doc Browser is where you'll find online help. Check out the
tutorials if your Python is rusty, or if you are unfamiliar with IPython, a version of interactive
python often used for scientific computing in Python. Your version of Canopy may also have
a "Training on Demand" option. Feel free to peruse it if it's available.

30

Canopy's Package Manager provides a fairly convenient way to manage your environments
packages and their dependencies. Open up the Package Manager to see which packages
are available in Canopy, which are currently installed, and if there are any that can be
updated. Throughout the course we'll be using some optional packages that you'll install
35   using the Package Manager. We'll be relying heavily on the **pandas** and **numpy** packages,
so you should check now to make sure they are installed, and if they aren't, you should install
them.   Other packages we'll be using in later sessions of thise course include **matplotlib** and
SQLAlchemy. You can install these now if you wish, or you can wait until you need them to
do so.

40

Do you have *sqlite3* installed on your computer? *sqlite3* is a widely used, public domain,
serverless SQL database engine that you can interact with from Python, from other computing
environments and languages, and from the command line. It's a small and easy-to-use
engine that's suitable for many kinds of applications. Linux and Mac distributions usually
45   include it. You may find it useful to have in this course. You can download it from
http://www.sqlite.org.

So now that you have Canopy installed (Did you try a "Hello World!" program with it?), here's a bit more about XYZ.   The XYZ Company is a retailer of consumer electronics and

50   appliances doing business through its network of 200 stores, its Web properties, and via it's cataloging and direct-mail campaigns.   Two documents about XYZ are provided for your perusal, a mock-up of XYZ's ecommerce site, and a one page overview of the firm. These documens are also used in PRED 450, the marketing analytics course.

55   In recent years XYZ has been investing more and more in its data analytics capabilities with the goal of becoming a more effective data-driven marketer. Like many firms, along the way in pursuing this objective, some things have gone well for XYZ, and other things, not so well. There has been some turn-over in its data analytics and database management team.   Some of the departed appear to have either perpetrated unnatural acts on XYZ's data assets, or to

60   have been just neglectful. (None of these individuals were MSPA graduates, of course.)